OPTIMIZATION OF OPTICAL AND RADAR SATELLITE DATA IN GOOGLE EARTH ENGINE FOR MONITORING OIL PALM CHANGES IN TROPICAL RIVER BASINS

ZENG JU

UNIVERSITI SAINS MALAYSIA

2023

OPTIMIZATION OF OPTICAL AND RADAR SATELLITE DATA IN GOOGLE EARTH ENGINE FOR MONITORING OIL PALM CHANGES IN TROPICAL RIVER BASINS

by

ZENG JU

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

September 2023

ACKNOWLEDGEMENT

My Ph.D. is a very rewarding period. I am really fortunate to have met professors and friends who have supported me in both my research and personal life.

First, I want to sincerely thank Associate Professor Dr. Tan Mou Leong, my main supervisor, and my co-supervisor, Prof. Dato. Dr. Narimah Samat, for their guidance in my Ph.D. research. They have always patiently guided me along the journey. When I first enrolled, I was confused and unclear what I wanted to study for my Ph.D. With Dr. Tan's help, I started to explore the latest research gaps and eventually defined the research directions successfully. He took great efforts to improve the writing of my research articles. He gave me praise when the first article published, and I still remember it. He gently advised me not to deny myself when the second article was rejected the work had simply not yet been met the right journal. His support, encouragement and constructive comments finally helped the second article get accepted and published in a Q1 journal. He always encouraged and supported us to attend conferences to know the latest development and expand networking. His extensive knowledge, advanced research skills and down-to-earth supervision style have set an outstanding example for me as an academician.

In addition, I would like to thank Miss Tew Yi Lin who has assisted in both my research and personal life. I remember how she shared me with a lot of study resources when I initially came in Malaysia. Thanks her for bringing us to purchase essential food and non-food items during the COVID-19 lockdown period. She has also guided me how to use the data collection software during the field trips.

I also express my gratitude to my family for their encouragement and support, especially my husband Wang Tao for accompanying me through all the challenges during my studies. When I was too busy writing my PhD thesis, he is the one who take care of my parents and two kids. When I had the idea of giving up my studies, his encouragement and understanding gave me strength to continue the journey.

Lastly, I would like to acknowledge the Ministry of Higher Education Malaysia for supporting the fieldwork and research data through the Long-term Research Grant Scheme (LRGS) project 2 with the grant number of LRGS/1/2020/UKM-USM/01/6/2.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

km	Kilometer
ha	hectare
RM	Malaysian ringgit
Mt	Million Tons
mm	Millimeter
˚C	Degree Celsius

LIST OF ABBREVIATIONS

AGB	Above ground biomass
ALOS	Advanced Land Observation Satellite
ANN	Artificial Neural Network
AOPD	Annual Oil Palm Area Dataset
ASM	Angular Second Moment
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer Global
AVG	Average
AVE	Average
AW3D30	ALOS World 3D-30m
C	Combination
CA	Consumer accuracy
CA- Markov	Cellular Automat Markov
CART	Classification and Regression Trees
CBERS	China–Brazil Earth Resources Satellite program
CODE-DE	Copernicus Data and Exploitation Platform-DE
CON	Contrast
COR	Correlation
CS	Component substitution
dB	Decibels
DEM	Digital Elevation Model
DIS	dissimilarity
DIF	Difference
DN	Digital numbers
DST	Daytime Surface Temperature

ENT	Entropy
ESA	European Space Agency
EVI	Enhanced Vegetation Index
EW	Extra Wide swath
FELDA	Federal Land Development Authority
FRSE	forest
GEE	Google earth engine
GEOS	Geostationary Operational Environmental Satellite
GIS	Geographic information system
GLCM	Grey-level co-occurrence matrix
GRD	Ground range detected
GSD	Ground sampling distances
HCV	High conservation value
HH	Horizontal-horizontal
HPC	High Performance Computing
HV	Horizontal-vertical
IDE	Interactive Development Environment
IDM	Inverse Differential Moment
IDL	Interface description language
IGSO	Improved grid search optimization
ILWRM	Integrated land and water resources management
IW	Interferometric Wide
JAXA	Japan Aerospace Exploration Agency
JM	Jeffries - Matusita
JRB	Johor River Basin
KEJORA	South East Johor Development Authority
KNN	K-Nearest Neighbors

LULC	Land-use and land-cover
MaD	Mahalanobis Distance
MD	Minimum Distance
MLC	Maximum likelihood classifier
MODIS	Moderate Resolution Imaging Spectroradiometer
MPC	Microsoft Planetary Computer
MRB	Muda River Basin
MS	Multispectral
MSD	Multi-scale decomposition
MSI	Multispectral imager
Mtry	Number of features
NASA	National Aeronautics and Space Administration
ND	Normalized difference
NDI	Normalized difference index
NDII	Normalized Difference Infrared Index
NDMI	Normalized Difference Moisture Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NEX	NASA Earth Exchange
NIR	Near-infrared
NN	Neural networks
Ntree	Number of trees
OA	Overall accuracy
OILP	oil palm
OLI	Operational Land Imager
PA	Producer accuracy
PALSAR	Phased Array L-band Synthetic Aperture

PAN	Panchromatic
POLSAR	Polarization Synthetic Aperture Radar
RF	Random Forest
RSPO	Roundtable on Sustainable Palm Oil
RUBR	Rubber
SAGA	Automated Geoscience Analysis
SAM	Spectral Angle Mapper
SAR	Synthetic Aperture Radar
SAVI	Soil adjusted vegetation index
SC	Supervised classification
SLC	Single-view composite
SM	Stripmap
SRBs	Spectral Reflectance Bands
SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machine
SWIR1	Short-wave infrared 1
TIRS	Thermal Infrared Sensor
TM	Thematic mapper
TOA	Top-of-atmosphere
USGS	United States Geological Survey
VAR	Variance
VIIRS	Visible Infrared Imaging Radiometer Suite
VH	Vertical-horizontal
VV	Vertical - Vertical
WATR	Water
WV	Wave

LIST OF APPENDICES

PENGOPTIMUMAN DATA SATELIT OPTIK DAN RADAR DALAM GOOGLE EARTH ENGINE UNTUK MEMANTAU PERUBAHAN KELAPA SAWIT DI LEMBANGAN SUNGAI TROPIKA

ABSTRAK

Pemetaan ladang kelapa sawit yang tepat adalah penting untuk merancang alaman pengurusan pertanian terbaik. *Google Earth Engine* (GEE), platform pengkomputeran berasaskan awan, membolehkan pengguna memproses imej satelit pelbagai sumber dengan lebih cepat dan berkesan. Sebenarnya, adalah sukar untuk membezakan kelapa sawit dengan tanaman lain hanya menggunakan satelit optik kerana isu litupan awan di kawasan tropika. Malangnya, hanya terdapat sedikit pemahaman saintifik tentang bagaimana pelbagai imej satelit dalam GEE boleh membantu dalam pemetaan ladang kelapa sawit. Justeru, kajian ini bertujuan untuk menentukan kombinasi optimum sumber terbuka data satelit optik dan radar untuk pemetaan ladang kelapa sawit di lembangan sungai tropika menggunakan Lembangan Sungai Muda (MRB) dan Lembangan Sungai Johor (JRB) sebagai tapak kajian. Pertama, dua pengelas pembelajaran mesin yang terdapat dalam GEE, hutan rawak (RF) dan mesin vektor sokongan (SVM), telah dibandingkan untuk mengenal pasti pengelas yang paling berkesan untuk pemetaan ladang kelapa sawit. Kemudian, lapan kombinasi data berbeza telah dibina daripada imej satelit dan indeks seperti C-band Sentinel-1, L-band PALSAR2, Landsat8, Sentinel-2, topografi, *Normalized Difference Vegetation Index* (NDVI), *Normalized Difference Water Index* (NDWI) dan lain-lain. Akhir sekali, gabungan data optimum digunakan untuk mengunjurkan taburan kelapa sawit pada masa hadapan dengan menggunakan metod CA-Markov. Penemuan menunjukkan bahawa RF mempunyai prestasi lebih baik daripada SVM dalam

pemetaan ladang kelapa sawit di kedua-dua lembangan sungai. Ketepatan peta gunaan dan liputan tanah yang dihasilkan selepas mengoptimumkan imej satelit adalah antara 93% hingga 95%, dimana Sentinel-1 dan Landsat8 manghasilkan keputusan pengelasan keseluruhan terbaik. Gabungan PALSAR2 dan Landsat8 menunjukkan pengelasan terbaik bagi ladang kelapa sawit, dengan nilai ketepatan pengeluar dan pengguna masing-masing sebanyak 91% dan 93%. Ketepatan pengelasan kelapa sawit boleh dipertingkatkan dengan menggabungkan imej radar C-band, namun keluasan ladang kelapa sawit adalah lebih rendah jika dibandingkan dengan imej L-band. Adalah penting untuk memilih kombinasi optimum untuk pemetaan kelapa sawit kerana empat kombinasi berasaskan radar menghasilkan kawasan kelapa sawit yang berbeza untuk MRB dan JRB pada 2020, masing-masing antara 6% hingga 12% dan 1% hingga 13%. Berbeza dengan ladang kelapa sawit di JRB, yang berkembang lebih perlahan sejak 2020 dan mula merosot selepas 2025, ladang kelapa sawit MRB akan berkembang secara konsisten dari 2020 hingga 2030. Penyelidikan ini boleh menjadi panduan untuk menambah baik pemetaan ladang kelapa sawit kos rendah dari angkasa di lembangan sungai tropika.

.

OPTIMIZATION OF OPTICAL AND RADAR SATELLITE DATA IN GOOGLE EARTH ENGINE FOR MONITORING OIL PALM CHANGES IN TROPICAL RIVER BASINS

ABSTRACT

Accurate mapping of oil palm plantations is crucial for planning agricultural best management practices. Google Earth Engine (GEE), a cloud-based computing platform, allowing users to process multi-source satellite images more quickly and effectively. In fact, it is difficult to distinguish oil palm from other crops using only optical satellites due to the issue of cloud cover in tropical regions. Unfortunately, there is only little scientific understanding about how various satellite images within GEE can be helpful for mapping oil palm plantations. Hence, this study aims to determine the optimal combination of open-source optical and radar satellite data for mapping oil palm plantations in tropical river basins using the Muda River Basin (MRB) and the Johor River Basin (JRB) as test sites. First, the two machine learning classifiers available in GEE, random forest (RF) and support vector machine (SVM), were compared to identify which is the most effective classifier for mapping oil palm plantations. Then, eight different data combinations have been constructed from the satellite images and indices such as C-band Sentinel-1, L-band PALSAR2, Landsat8, Sentinel-2, topographic, Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), etc. Lastly, the optimal data combination was employed to project future oil palm distribution using the CA-Markov approach. The findings demonstrate that RF outperformed SVM in mapping oil palm plantations in both river basins. The accuracy of land use and land cover maps generated after optimizing the satellite images ranging from 93% to 95%, with

Sentinel-1 and Landsat8 provided the best overall classification. The combination of PALSAR2 and Landsat8 demonstrated the best classification for oil palm plantations, with the computed producer and consumer accuracy values of 91% and 93%, respectively. The accuracy of oil palm classification can be improved by combining C-band radar images, however the area of oil palm plantations is underestimated when compared to L-band images. It is important to select the optimal combination for oil palm mapping since the four radar-based combinations resulted different oil palm areas for MRB and JRB in 2020, ranging from 6% to 12% and 1% to 13%, respectively. In contrast to the oil palm plantations of JRB, which expand more slowly since 2020 and starts to decline after 2025, the oil palm plantations of MRB will expand consistently from 2020 to 2030. This research could serve as a guide to improve the low-cost mapping of oil palm plantations from space in tropical river basins.

CHAPTER 1

INTRODUCTION

1.1 Research Background

Oil palm is one of the most rapidly growing agricultural land uses, while the majority of crops remain relatively steady in their production area. The oil palm is commonly referred as a "golden crop" due to its high-yielding, which is more than that of any other oil crop (Mohd Najib et al., 2020). As among the world's largest agricultural plantations, oil palm plantations have a substantial impact on ecology, environment, and economy (Alam et al., 2009). Driven by the high worldwide demand for oil palm products and the potential for future growth, if not properly monitored, uncontrolled oil palm activities can lead to deforestation with serious negative environmental impacts (Shaharum et al., 2020). Therefore, quantifying and predicting the dynamics and trajectory of oil palm planting areas and properly managing and monitoring oil palm planting activities are crucial for the development of sustainable oil palm management strategies.

Malaysia has a remarkable tropical monsoon and tropical rainforest climate with abundant precipitation year-round. March to June and October to February are rainy seasons, with an average annual rainfall between 2500 mm and 5080 mm. The annual average temperature changes little throughout the year, ranging from 20 ℃ to 30 ℃. These climatic characteristics are ideal for oil palm planting and growth (Corley & Tinker, 2008). The rapid growth of oil palm under the government's diversification program in the early 1960s greatly reduced Malaysia's dependence on rubber and tin (Awalludin et al., 2015). Oil palm cultivation in Malaysia has increased rapidly, almost

exponentially, during the previous three decades, according to the United Nations Food and Agricultural Organization.

The palm oil industry has grown to be an important component of Malaysia's economic development, which as of 2012 employed about 490,000 people (Michael, 2012). Following Indonesia, Malaysia is the second-largest producer of oil palm in the world. Over 5.6 million hectares of oil palm estates and more than 400 palm oil-producing mills were present in Malaysia in 2015 (MPOB, 2020). Exports of oil palm products are more than $30 billion USD annually (FAO, 1997). In 2020, Malaysia exported about 17.40 Mt of oil palm to other countries, bringing in about US$15 billion (MPOB, 2020). Therefore, Malaysia makes huge profits when it exports oil palm products to other countries (Figure 1.1).

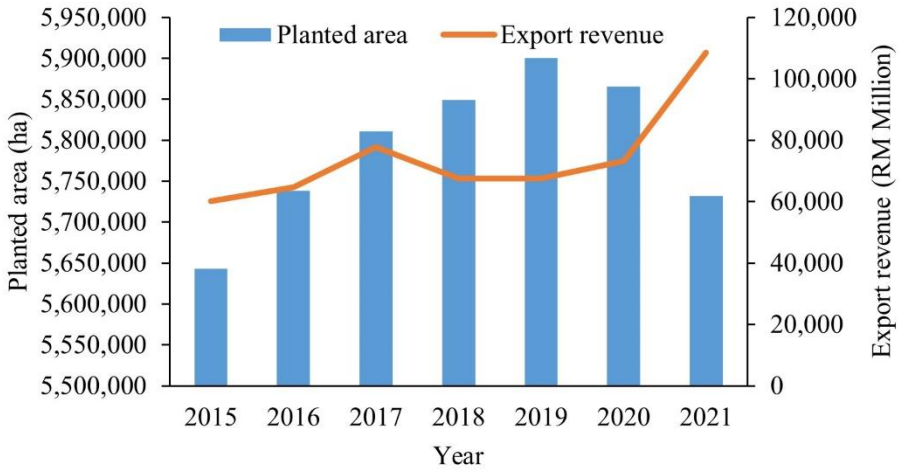

Figure 1.1 Oil palm planted area and export revenue in Malaysia from 2015 to 2021 (MPOB, 2020).

Although the oil palm sector is a major contributor to Malaysia's GDP, but the rapid expansion of oil palm plantations has resulted significant deforestation. The conversion of forest to oil palm is considered as a major threat to biodiversity in addition to contribute to global warming and climate change. Currently, 16.4% of Malaysia's total land area is now used to cultivate oil palm (Mohd Najib et al., 2020).

In Malaysia, the conversion of agriculture land and the clearing of forests in Malaysia have contributed to an expansion of more than 50% in oil palm plantations than before (Koh & Wilcove, 2008). The ecosystem has been severely harmed by this expansion, including the loss of flora and fauna (Sheil et al., 2009).

The expansion of oil palm plantations within a river basin could result a substantial impact on the availability and quality of water resources (Kang & Kanniah, 2022; Tan et al., 2021). Accurate maps of the oil palm distribution for multiple years are essential to comprehend patterns in oil palm plantation expansion for agricultural best management practices planning (Danylo et al., 2021). Fieldwork-based traditional methods are characterized by labor and time inefficiencies. Using remote sensing technology to collect, monitor and track the changes in oil palm distribution is a more practical approach (Gong et al., 2012).

In recent years, satellite remote sensing has been extensively used for mapping and monitoring of the distributions of oil palms (Chong et al., 2017). The effectiveness of oil palm classification is constrained by frequent cloud cover in the tropics, so it is challenging to map of oil palm distribution. Initially, mostly low-resolution (250–1000 m) MODIS data and Landsat TM remote sensing imagery were mainly utilized to assess the net primary productivity (Cracknell et al., 2013; Tan et al., 2011) and age of oil palm plantations (McMorrow, 2001; Vadivelu et al., 2014).

The problem of poor quality optical data caused by cloud cover may be efficiently resolved by combining radar images (Sarzynski et al., 2020). Synthetic Aperture Radar (SAR) and optical remote sensing data have been integrated in several studies to provide oil palm mapping with complementing characteristics (Cheng et al.,

2016; De Alban et al., 2018; Mohd Najib et al., 2020; Sarzynski et al., 2020), showing that the combined imagery improves the accuracy of oil palm classification.

Data fusion involves gathering relevant information from various data sources and storing it in a small number of data repositories, usually just a single one (Himeur et al., 2022; Lee et al., 2016). It is the process of combining many data sources to provide more consistent, accurate, and relevant information than that offered by any one data source alone. In order to obtain more accurate information, multi-source remote sensing image fusion has emerged a research hotspot in the field of remote sensing. Combinations of image information obtained by different sensors help to accomplish more accurate and comprehensive remote sensing earth observation (Zhang, 2010).

Researchers in the geospatial data science and remote sensing domains have expressed tremendous interest in the development of GEE (Tamiminia et al., 2020). According to a GEE research of land cover change in Singapore, the impact of monsoonal cycles on forest cover is greater than anthropogenic factors (Sidhu et al., 2018). Using Landsat and GEE, Oliphant et al. (2019) mapped farmland across Southeast Asia and Northeast Asia. Rudiyanto et al. (2019) utilized the Sentinel-1 data available from the GEE platform to automatically map and monitor Southeast Asian rice extent, cropping patterns, and growth phases in near-real-time. In a South African semi-arid environment, Gxokwe et al. (2022) employed GEE to study seasonal wetlands changes.

Oil palm mapping and monitoring are made easy by GEE's fast computing capacity and extensive multi-source data (Amani et al., 2020). Shaharum et al. (2020) employed GEE to perform oil palm mapping in Peninsular Malaysia, with the

classification accuracy of more than 80%. Similarly, Li et al. (2020b) also utilized GEE to analyze the temporal and spatial dynamic of the oil palm patterns as well as their driving forces in Malaysia between 2000 and 2018.

The use of multi-source remote sensing images for GEE-based oil palm classification has attracted the attention by a few researchers. For example, Sarzynski et al. (2020) employed GEE to combine radar and optical imagery for creating a comprehensive oil palm map of Sumatra, Indonesia, whereas Poortinga et al. (2019) fused images from Landsat 8, Sentinel-2, and Sentinel-1 satellites to map Burmese oil palm accurately. By utilizing Sentinel-1 and Sentinel-2 data within the GEE platform Abramowitz et al. (2023) improved land cover mapping in Ghana with a focus on distinguishing between oil palm plantations and natural forests.

1.2 Problem Statement

Google Earth Engine (GEE) is a cloud-based computing platform developed by Google, providing massive archive of satellite and geospatial data, simple visualization operations, and powerful analytical capabilities. The platform offers satellite data from multiple sources at various scales, including Sentinel-1/2, Landsat TM (Thematic mapper)/OLI (Operational Land Imager), and MODIS. More than 5 million satellite images and 200 public datasets are included in the data archive, which has a petabyte-level capacity (Gorelick et al., 2017). GEE-based platforms made it simpler to conduct large-scale and global-level analysis via the cloud computing platform as compared to the traditional approach, which required users to download and process the satellite images (Tamiminia et al., 2020). Researchers have utilized GEE to track changes in different environments and applications, including forest (Chen et al., 2017), aridity (Rembold et al., 2019), surface water (Pickens et al., 2020),

flood (Coltin et al., 2016; Tew et al., 2022a), crop (Dong et al., 2016) and aquaculture pond (Tew et al., 2022b). When creating land-use and land-cover (LULC) maps across wide areas or long periods, temporal aggregation is a popular technique for addressing data gaps and disparities in the available satellite images (Gebhardt et al., 2014; Verhegghen et al., 2016; Winsvold et al., 2016). The GEE platform can quickly process and aggregate hundreds of multi-phase satellite images. However, the classification accuracy may reduce from insufficient satellite images or poor cloud removal algorithms in GEE, (Phan et al., 2020).

Effective image classification algorithms are necessary to extract accurate information about oil palm from satellite images. Some researchers have examined the reliability of random forest (RF) algorithms within GEE to classify land uses based on optical images for large regions (Monsalve-Tellez et al., 2022). While, another researchers have explored the potential of integrating optical and SAR images to enhance the precision of classification (Carolita et al., 2019). In numerous studies, SVM has shown to execute classification tasks better than common classification methods (Cheng et al., 2016; Nooni et al., 2014; Toh et al., 2019). Notably, oil palm plantations have been successfully mapped using SVM, which effectively handles high dimensional data without reducing its dimensionality. On the other hand, RF offers high classification performance and anti-noise ability (Briem et al., 2002; Jin et al., 2018). In fact, RF is the algorithm most commonly employed in land cover analysis as evidenced by a number of studies (De Alban et al., 2018; Goldblatt et al., 2018; Sarzynski et al., 2020). The direct use of a single algorithm limits the research of understanding the effectiveness of other classification algorithms (Shaharum et al., 2020). Therefore, it is quite interesting to understand how well RF and SVM work in the context of mapping oil palm plantations.

Since the 1990s, oil palm monitoring has made use of remote sensing technologies (Xu et al., 2020). Large amounts of cloud cover in tropical regions limit the capacity of optical satellites to track the oil palm changes continuously. During a given timeframe, fusion products frequently provide more information than single-band images. Therefore, image data fusion or combination technologies can be utilized to address or partially mitigate the spatial and temporal limitations of a single sensor. The fusion of MODIS and Landsat data has been extensively employed. While, data fusion technologies that combine optical and SAR images are also utilized for forest mapping and surface water detection (Poortinga et al., 2019), but attention to mapping oil palm plantations is still limited.

Numerous researchers have created oil palm distribution maps by utilizing multi-source remote sensing data of various temporal and spatial resolutions (Poortinga et al., 2019; Sarzynski et al., 2020; Xu et al., 2020). The integration of optical images with SAR images has become more popular as a highly effective approach to improve the precision of oil palm mapping in tropical regions (Mohd Najib et al., 2020; Oon et al., 2019b; Poortinga et al., 2019; Sarzynski et al., 2020). This overcomes the issue of limited images due to cloud cover in tropical regions. For example, Cheng et al. (2016) have used Landsat and Phased Array L-band Synthetic Aperture (PALSAR) data to map oil palm plantations in Malaysia, focusing on assessing the impact of various classifiers, locations, and evaluation techniques. Mohd Najib et al. (2020) created an oil palm map for Malaysia using Landsat and Advanced Land Observation Satellite (ALOS) images. Their findings revealed that the estimated area of oil palm plantations was slightly larger than what was stated in the official statistic data. According to Oon et al. (2019b), L- band and C-band radar images performed better than other sensors in tropical regions, especially when it came to

differentiating between industrial and smallholder oil palm plantations in the peatland region of Peninsular Malaysia. The effects of various fusions of optical and radar images on the mapping of oil palms, however, have received less attention in earlier studies. In reality, using information from multiple satellite sensors is necessary for comprehensive monitoring of the distribution of oil palm plantations (Gutiérrez-Vélez & DeFries, 2013).

Utilizing open-source satellite data and the computational power of GEE has made it possible to acquire more accurate information on the extent of oil palm cultivation in developing countries (Amani et al., 2020; Gorelick et al., 2017), while spending less cost (Lee et al., 2016; Puttinaovarat & Horkaew, 2018; Shaharum et al., 2020). Some researchers employed GEE to integrate images from many sources into a single image to create a map showing the distribution of oil palms. For example, Sarzynski et al. (2020) employed GEE to fuse radar and optical images to analyze the spatial distribution of oil palm on the island of Sumatra. The results shown that using both optical and radar data together was better to using either optical data or radar data solely. Some researchers employ satellite images at various times to analyze the long-term changes in oil palm in order to overcome the problem of incomplete data. Using the Landsat and L-Band SAR data, De Alban et al. (2018) mapped tropical landscapes for the classification of land cover and change detection. Several researchers combined multiple sources image data to gather additional information about oil palm plantations based on the distinctive features on the ground. For instance, Danylo et al. (2021) employed Sentinel-1 data to obtain the location of oil palm plantations, whereas the determination of oil palm plantation age was accomplished by the utilization of Landsat imagery. There is no single form of data can be used to classify oil palm in all regions, since various factors and goals can alter the accuracy of the results (Torbick

et al., 2016). Therefore, it is necessary to assess the optimal combination of multi-source data to produce more accurate oil palm maps. However, there are still relatively few studies on identifying the optimal way to fuse optical and radar satellite data within the GEE platform, particularly for oil palm mapping,

Land-use and land-cover (LULC) change is one of the main causes of the shift in the hydrological cycle (Tan et al., 2015). Hence, implementing integrated land and water resources management (ILWRM) requires a rigorous assessment of LULC changes (Badjana et al., 2015). It will be challenging to manage land and water resources sustainably if the LULC data is of poor quality (Zurqani et al., 2018). Therefore, high-precision LULC maps are needed in order to reliably simulate the influence of LULC changes on water resources in the basin.

CA-Markov is an effective tool for simulating and detecting LULC change because it takes into account the spatial and temporal components (Hyandye & Martz, 2017). The hybrid model is based on the CA architecture by utilizing the excellent expressive ability of the CA model and the predictive power of the Markov model. LULC is predicted by combining the two models, multi-standard and multi-objective land allocation principles (Zhou et al., 2020), producing more accurate prediction results. The model has been widely used in predicting future LULC changes in the tropics due to the simplicity, flexibility, and capability of integrating GIS and remotely sensed data (Noszczyk, 2019; Rahnama, 2021),

The CA-Markov model was first used by Nourqolipour et al. (2011) to simulate oil palm changes, but the study lacked information on the environmental and socio-economic aspects of oil palm development. In order to simulate the expansion of oil palm plantations in Kuala Leng Yueh, Malaysia, Nourqolipour et al. (2015) used

constraints and nine suitability factors to construct assessment criteria that were integrated with CA-Markov. The results showed that the spatial expansion of oil palm plantations. Camara et al. (2020) used the CA-Markov model to project future changes in land use in the Selangor River Basin (SRB), Malaysia. The findings demonstrated the model was successfully in projecting LULC changes in 2024 and 2033, but the study only focused at a river basin that dominated by natural forests. Although it is possible to predict the distributions of oil palm using the CA-Markov model, but there is little study on past related to the future projections of oil palm distribution in tropical river basins, so the applicable remains unclear.

Muda River Basin (MRB) and Johor River Basin (JRB) are important sources of freshwater supply in Malaysia and Singapore. Therefore, it is vital to have a comprehensive understanding of the impact of LULC changes, mainly the increase of oil palm plantations, on the climate and environment in these regions. Tan et al. (2021) examined the impacts of oil palm expansion on MRB water balance using improved European Space Agency (ESA) land cover products. Whereas, the impacts of LULC on the river morphology of the JRB in the southern part of Peninsular Malaysia were examined by Kang and Kanniah (2022) using GEE. Little study has been done on the identification and extraction oil palm information within these two basins, Nevertheless, the evaluation of the synergistic effect of several satellite image types inside GEE for oil palm mapping is currently limited in tropical basins, particularly in Malaysia.

1.3 Research Questions

1. How different classification methods in GEE impact the accuracy of the oil palm mapping?

2. Which combinations of multi-source remote sensing images are effective for mapping oil palm in tropical river basins?

3. What are the changes of historical and future oil palm plantations in tropical river basins?

1.4 Research Objectives

The objective of this study is to develop a framework for optimizing the optical and radar satellite data in tropical basins, specifically for the purpose of monitoring oil palm plantations. This study specifically focused on two tropical basins, the MRB and the JRB, with the following specific objectives:

1. To evaluate the accuracy of random forest (RF) and support vector machine (SVM) classifiers for mapping oil palm plantations in tropical river basins.

2. To determine the optimal combination of multi-source optical and radar satellite data for mapping oil palm plantations using the GEE platform.

3.To project future oil palm changes in in tropical river basins for the year 2030 using the CA-Markov model.

1.5 Significance of the Study

A key factor in getting accurate satellite image classification for LULC mapping is the selection of classifiers. Currently, SVM and RF are widely recognized as the two most often employed classification methods in machine learning classifiers. Both methods can achieve good classification accuracy in oil palm mapping. However, it should be considered which method can better adapt to the classification of oil palm in watersheds with poor image quality in tropical areas, with similar ground objects. This study can serve as a reference for scholars to select the best classifier for similar studies before carrying LUCL or oil palm classification in this regions.

The accuracy of oil palm classification results can be influenced by several circumstances and purposes, making it unsuitable to rely on a single data type for all oil palm regions. Therefore, different combinations of multi-source remote sensing images help to understand the differences in the extraction of oil palm, so that the optimal data combination can be used to create oil palm maps for hydrological or agricultural yield modelling purposes. This study provides a reference for choosing a reasonable and rapid detection method for mapping oil palm plantations. The open source data of the GEE platform can be fully utilized to construct the best oil palm maps for the research area. With the GEE-based oil palm plantations mapping framework, users no longer need to download the data to a local machine for stand-alone processing. Compared with the traditional data processing, this proposed GEE-based framework greatly improves the operating efficiency and saves local storage space. The framework can also increase the precision of LULC classification and oil palm monitoring in tropical river basins. These findings may help researchers in other

tropical nations create more accurate maps of oil palm distribution, providing oil palm managers with more reasonable data for estate planning and management.

While oil palm brings significant economic benefits, rapid expansion can also lead to adverse environmental and social impacts. Accurate LULC data serves as the foundation for pertinent scientific studies such as analyses of the carbon cycle, hydrological modeling and soil analysis. Additionally, the oil palm plantation maps are fundamental knowledge for local land agencies to plan future management strategies.

Both MRB and JRB are important sources of water supply in Malaysia. Dramatic changes in LULC within the basins may largely impact on local hydrological components such as runoff, water availability, streamflow, groundwater flow and evapotranspiration. This study employs the CA-Markov model to predicted future LULC maps of two river basins based on the theoretical scenarios. The LULC maps that are reasonably predicted can be utilized as important input data for hydrological models and auxiliary data for future basin planning, management, and sustainable development. The findings are helpful for further research in the areas of ecosystem protection and management, sustainable land use planning, mitigation of natural catastrophes such as floods, and ensuring water safety (Kang & Kanniah, 2022).

1.6 Scope of the Study

The main goal of this study is to better understand the capability of multi-source remote sensing images for oil palm classification in tropical river basins, focusing on the MRB and the JRB. The distribution of LULC of these two basins is quite diverse, with rubber and rice being most distributed in the MRB, while oil palm

predominating in the JRB. Oil palm crops in these two basins has changed significantly during the previous two decades, resulting in the expansion of oil palm plantations emerging as a prominent land use category within these basins. Using two river basins in the north and south parts of Peninsular Malaysia with vastly different land cover types as research objects, the potential and applicability of the optical and SAR images integration can be better discovered. Basically, the scope of this study is divided into the following three parts:

The first part constructed different combinations of optical and radar images for LULC classification of MRB and JRB in 2020 based on the multi-source data from the GEE platform. High-resolution images are not always freely available and difficult to compatible with the final user due to the different resolutions offered by different satellite platforms. All public data in the GEE platform is free and open to all non-commercial application users, and the algorithm embedded in the platform can easily process the selected data. The GEE cloud platform is an online webpage, and users can run multiple webpages at one time, and conduct regional or global-scale research at any time and place. These data include Landsat series data, Sentinel series data, Advanced Land Observation Satellite 2 Phased Array L-Band Synthetic Aperture Radar 2, and SRTM (Shuttle Radar Topography Mission) data. Using field-collected data and high-definition images, the obtained classification map was verified and analyzed from two aspects: statistical value and visual analysis.

The second part compared and analyzed the classification accuracy of different combinations and optimized the characteristic parameters of optical and radar image combination to enhance the discriminatory capability between oil palm and other potentially confounding crops (rubber). In the specific optimization process, the

eigenvalues of the radar image are optimized mainly according to the reflection degree of different objects on the image and the contribution degree to the parameters in the random forest classifier. Then compare and analyze the classification diagrams before and after optimization. Classification methods such as RF, SVM, MD, and CART are commonly used supervised classification methods in oil palm or LULC classification. This study utilized the SVM and RF classifiers available on the GEE platform and analyzed their capabilities in terms of classification accuracy.

The third part analyzed the changes of LULC in the two river basins from 2000 to 2020. This analysis primarily focuses on the change and expansion features of oil palm cultivation within the watershed, utilizing the optimal classification map. To quantify changes in LULC, the classification maps from the years 2000, 2010, and 2020 were selected. Subsequently, the land use transition matrix was employed to examine the modifications and expansion patterns of oil palm cultivation over this two-decade period. The CA-Markov model was employed to project land use trends in the MRB and JRB watersheds from 2020 to 2030 under the theoretical scenario, especially the expansion trend of oil palm.

1.7 Organization of the Chapters

The thesis is divided into six chapters, with each chapter comprising various components. The subsequent sections provide a concise overview of the components within each chapter.

Chapter 1 presents the background of the study, problem statement, research questions, research objectives, significance of the study, scope of the study and organization of the chapters.

Chapter 2 first introduces oil palm, including its characteristics and the development of oil palm in Malaysia. Then, the application of remote sensing to oil palm classification, mapping, and monitoring is reviewed. Some commonly used LULC prediction models and methods are summarized in detail, especially the CA-Markov model. Then, the sources of multiple remote sensing data that commonly used for oil palm mapping are summarized. Next, the literature focuses mostly on widely applied data fusion approaches of multi-source remote sensing data as well as feature optimization. The classification methods employed for remote sensing oil palm mapping, popular cloud-based computing platforms, and the current state of GEE-based oil palm research are summarized.

The research methods and data required for this study are described in Chapter 3, along with the geographical descriptions of two river basins. The chapter also describes the data processing platform, multi-dimensional feature datasets, creation of multi-source data combinations, classifier selection, multi-source indices optimization, accuracy assessment, LULC change detection, and the CA-Markov model.

The main findings of the study are primarily presented in Chapter 4, which also include the effects of various classifiers and image combinations in the two river basins. Then, the reliability of the LULC classification results, before and after image optimization, RF variable importance analysis, oil palm expansion over the past 20 years, and projections for the years up to 2030. Chapter 5 mostly discusses the findings of Chapter 4. Lastly, the research overview and major conclusions are presented in Chapter 6, along with suggestions for future directions.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Accurate land use data serve as the foundation for scientific studies such as analysis of carbon cycle, hydrological modelling, and assessments of soil degradation. Additionally, these data are crucial for local land management agencies in the planning and managing LULC. Likewise, LULC change is also negatively affecting the Earth's surface in a variety of ways, including terrestrial ecosystems, water balance, biodiversity, and climate (Salazar et al., 2015; Sterling et al., 2012). Hence, this chapter reviews and discusses the theoretical and other findings that serve as the foundation for this study. Moreover, the relevant technology and procedures for oil palm monitoring in tropical river basins are comprehensively reviewed.

2.2 Oil Palm

Oil palm is a single-leaf perennial plant belonging to the oil palm genus and palm family (Hai, 2002). Palm oil, one of the world's three major vegetable oils, stands out as a very prolific crop within the vegetable oil industry. Oil palm is commonly referred to as the "King of Oil in the World" because it produces an average of 3.5 tons of palm oil per hectare annually (Li et al., 2015), which is 2-3 times that of coconuts and 7-8 times that of peanuts (Sheil et al., 2009). Additionally, it is also the type pf vegetation oil that is produced, consumed and traded the most globally.

The worldwide planted area of oil palm has increased significantly over the past several decades, rising from 3.6 million hectares in 1961 to 28 million hectares in 2020 (FAOSTAT, 2020), making it one of the fastest-growing tropical crops. While

oil palm production is crucial for local and regional economies, including rural development, it has also had negative social and environmental effects, particularly in Southeast Asia (Sheil et al., 2009). These negative effects include accelerated deforestation and related detriments to biodiversity and ecosystem services (Lee et al., 2016).

2.2.1 Oil Palm Properties

Oil palm is originated from western Africa, initially expanded from Sierra Leone, Liberia, Ghana, and Cameroon all the way to the Congo and near the equator (Shuit et al., 2009; Sowunmi, 1999). It is believed that the oil palm industry began in Southeast Asia, where it was first to be grown in large quantities in the 1970s and soon became a significant commercial crop in countries such as Malaysia, Indonesia, and Thailand (Sheil et al., 2009).

The oil palm can be distinguished by its single-stemmed trunk and discernible crown (Figure 2.1). Eight fronds make up a rank and are arranged in succession as they spiral outward from the stem's apex. Viewed from above, it looks like an eight-pointed star (Chong et al., 2017). As an industry standard to enhance productivity and allow for the most sunlight penetration, the most suitable planting pattern for oil palms is a triangular pattern with a specified spacing of 9 meters (Basiron, 2007).

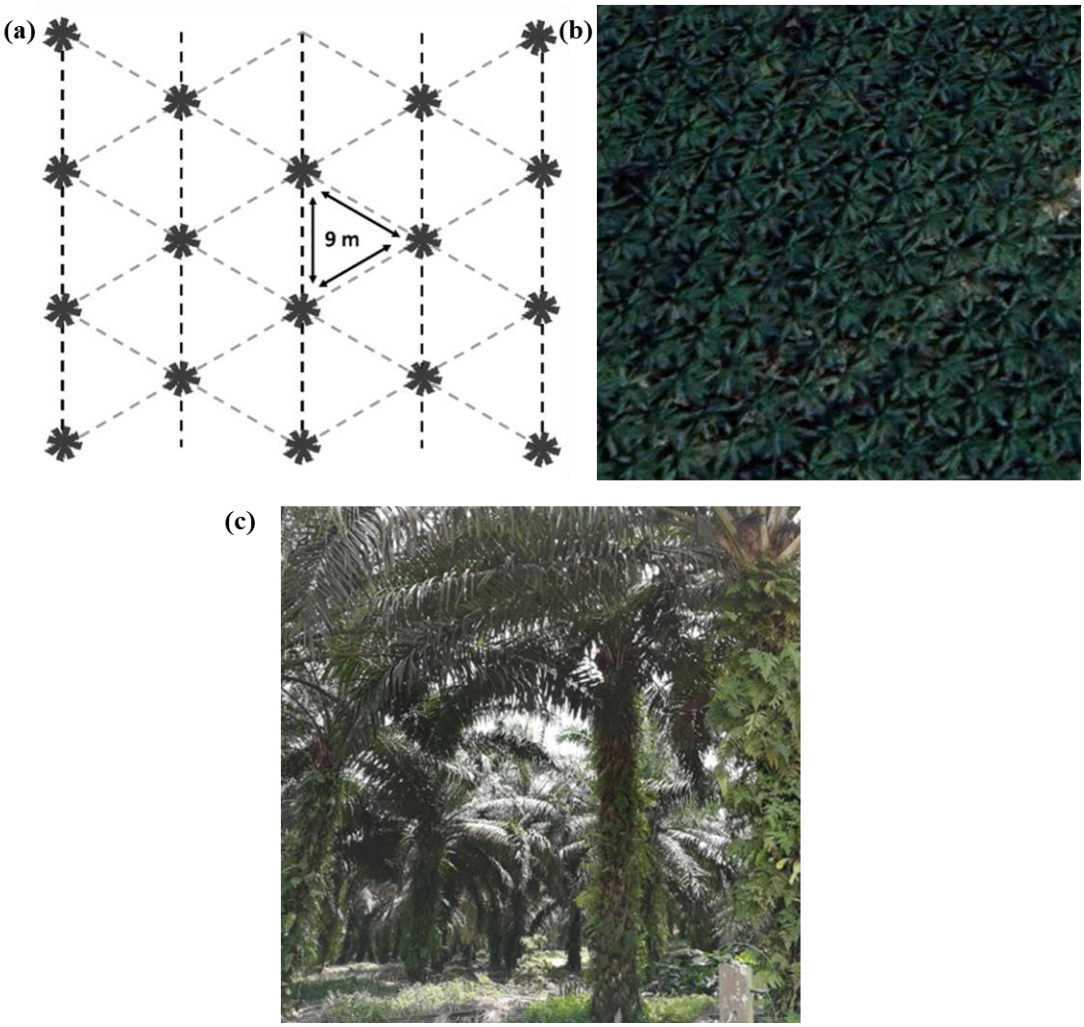

Figure 2.1 (a) Planting patterns of oil palm plantations (Chong et al., 2017); (b) oil palm plantations in JRB as seen from (b) Google Earth Pro; and (c) field visit in September 2020.

Oil palm needs a favorable climate system to grow nicely. Four main climatic parameters that support the growth of oil palm such as annual average temperature, average temperature during the coldest month of the year, annual precipitation, and the number of months with precipitation less than 100 mm (Pirker et al., 2016). The optimum temperature range for oil palm growth is between 24 and 28°C, with the coldest month of the year's average temperature being above 15°C. The annual rainfall should range from 2000 to 2500 mm, with a minimum monthly rainfall of 100 mm, which is conducive to the growth of oil palm plantations (Corley & Tinker, 2008).

The growth of oil palm does not have high requirements on the chemical and physical qualities of the soil, but when the water supply is insufficient, the growth of oil palm is easily affected. Therefore, considering the sensitivity of the water supply, the nutritional condition and water holding capacity of the soil should be considered when selecting the soil for planting oil palm. Oil palm can obtain comparable yields in soils with a high clay concentration as well as in loam and silt-dominated soils (Pirker et al., 2016).

Oil palm planting is limited by terrain because steep slopes increase the cost of planting, maintaining, and harvesting. Shallow soil and weak surface runoff on slopes limit the oil palm planting as well. Oil palm plantations can also be cultivated effectively on slopes up to 16°, however, 0 to 4° is the slope range that is optimum for their growth. According to conventional perception, the slope cannot be steeper than 25°. In addition, oil palm growth is affected by altitude, mainly because altitude is closely related to temperature and slope tendency (Pirker et al., 2016).

Therefore, oil palms are suitable for growing in tropical climates with high precipitation rates, high solar radiation, and temperatures of 24–32 ℃ (Corley & Tinker, 2008). According to the suitability of the climate, the land suitable for planting oil palm is predominantly located in 12 tropical nations, covering about 1.37 billion hectares (Pirker et al., 2016). Figure 2.2 depicts the global oil palm suitability map and the enlarged three oil palm important locations such as Amazon, the Central African coast, and Borneo Island.

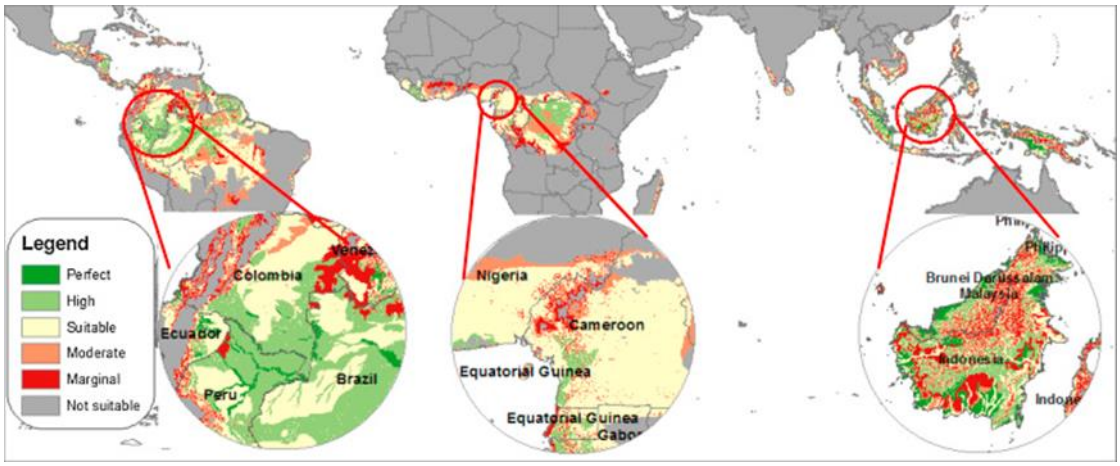

Figure 2.2 World oil palm suitability map with three focal areas enlarged in the Central African coast, the Amazon region, and the Borneo Island (Pirker et al., 2016).

2.2.2 Oil Palm in Malaysia

In 1971, a Frenchman called Henri Fauconier established the first oil palm estate in Selangor that was utilized for commercial purposes (Sumathi et al., 2008). Since then, oil palm farming has been practices in Malaysia. Following more than a century of growth, Malaysia is now ranks second globally in terms of both palm oil production and exports (Tang & Al Qahtani, 2019).

The Roundtable on Sustainable Palm Oil (RSPO) was founded in Malaysia to govern the sustainability of the oil palm plantation and industry (Tan et al., 2009). Furthermore, oil palm has been identified as a valuable renewable energy source, in addition to its principal use as a source of edible oil (Sumathi et al., 2008). The explosive growth of the oil palm industry is contributed by the increasing global demand for palm oil and the agricultural diversification initiative introduced by the Malaysian government in the early 1960s.

The total area of oil palm plantations in Malaysia was decreased by 0.6% from 5.9 million hectares in 2019 to 5.87 million hectares in 2020. With 1.59 million hectares or 27% of all oil palm plantations in Malaysia, Sarawak is the state with the

most oil palm plantations. Sabah is in second place with 1.54 million hectares, or 26.3% of the total oil palm plantation area. 2.74 million hectares, or 46.7%, of Malaysia's total oil palm plantation area is on the Peninsular Malaysia (MPOB, 2020). Johor has the largest oil palm cultivation area in Peninsular Malaysia, which is followed by Kedah, Kelantan, Melaka, and Negeri Sembilan.

While the oil palm industry's rapid growth has brought significant economic advantage, it has also led to a number of environmental issues, including the loss of tropical rainforests, the extinction of species, and the emission of greenhouse gases. To address these issues, Fitzherbert et al. (2008) investigated the effect of oil palm on biodiversity, Sumathi et al. (2008) described Malaysia's oil palm industry's current status to promote sustainable and renewable energy. Tang and Al Qahtani (2019) performed a sustainability analysis on oil palm plantations in Malaysia. Hence, reliable oil palm plantation maps are required to support planning and monitoring of oil palm growth (Cheng et al., 2016).

2.3 Remote Sensing

Remote sensing is defined as a long-range technology for obtaining information about the surface of the Earth surface continuously without physical contact (Atzberger, 2013; Pandey & Sharma, 2021). It is an advanced technique for tracking the growth and development of oil palm, which may be useful for analysing the impacts of oil palm plantations on environment and for selecting effective management strategies. Major applications of remote sensing in the oil palm industry include monitoring of oil palm plantation, tree counting, age estimation, above-ground biomass (AGB) and carbon production, pest and disease detection, and yield estimation (Asming et al., 2022).

2.3.1 LULC Classification

LULC monitoring is the identification and classification of samples of surface objects in remotely sensed images using classification methods. In remote sensing image classification, a number of variables are utilized to define the characteristics of the objects to be identified, and these variables are generally referred as image features or image knowledge (Li et al., 2017a).

Image feature extraction is the basis of remote sensing classification. An effective feature library is crucial to express the classified objects and final outputs. In addition, spectral signal, vegetation index, image variation form, texture, contextual information, multi-temporal images, multi-sensor satellites and other auxiliary data are a variety of variables that used in image interpretation.

The original feature variables used for computer image interpretation are the band images themselves, and the band images can be spatially operated to obtain variable features like texture and context. All of the above features are referred to as original features. The feature formation typically consists of the basic features produced by measuring the identified object with instruments or sensors. Figure 2.3 shows the LULC classification procedure for remote sensing images.

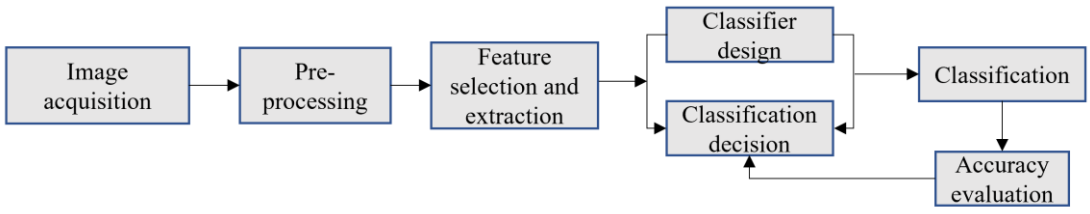

Figure 2.3 LULC classification workflow for remote sensing images.

2.3.2 Oil Palm Mapping

The rapid growth of oil palm in tropical regions has both positive and negative effects on the ecosystem. Accurate oil palm mapping can help to examine these effects. To evaluate the long-term effects of oil palm, it is important to regularly update the oil palm map. Hence, simple methods for mapping oil map plantations effectively and rapidly should be developed. Some researchers have exclusively studied SAR imagery for oil palm mapping (Cheng et al., 2018b; Li et al., 2015; Oon et al., 2019b), while others have combined optical and radar imagery (Cheng et al., 2016; Descals et al., 2021; Gutiérrez-Vélez & DeFries, 2013; Nomura et al., 2019; Pohl, 2014; Sarzynski et al., 2020; Vaglio Laurin et al., 2013; Xu et al., 2020).

Many researchers have mapped oil palm for large areas (Cheng et al., 2018a; Descals et al., 2021; Oon et al., 2019b; Rodríguez et al., 2021; Xu et al., 2020), with a focus on identifying oil palm from background LULC. Cheng et al. (2018a) mapped oil palm cultivation in the top 15 countries in the world in 2016 based on 100 m resolution PALSAR2 images and high-resolution data sourced from Google Earth, with the help of supervised classification and visual interpretation techniques. Descals et al. (2021) used deep learning and Sentinel-1/2 satellite data to create a 10 m-resolution global oil palm map of smallholder and industrial closed-canopy oil palm plantations in 2019. Oon et al. (2019b) compared the differences between radar remote sensing signatures of industrial and smallholder oil palm plantations, two distinct oil palm landscapes. The backscatter intensities of the ALOS-2 PALSAR2 L-band and Sentinel-1 C-band SAR were used to detect these features. For example, Xu et al. (2020) used satellite data from ALOS-2/PALSAR2 and MODIS to map the Annual Oil Palm Area Dataset (AOPD) at 100 m resolution for Malaysia and Indonesia from 2001 to 2016.